

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



TRABAJO FIN DE MÁSTER

**Desarrollo de una base de datos para la
clasificación de fármacos basada en el
proyecto Ontología Génica y KEGG drug
con aplicación en reposicionamiento de
fármacos**

**Máster Universitario en Bioinformática y Biología
Computacional**

Autora: BENÍTEZ DORTA, Raquel

**Tutores: AL-SHAHROUR NÚÑEZ, Fátima;
TEJERO FRANCO, Héctor**

Ponente: DÍAZ URIARTE, Ramón

**Unidad de Bioinformática
Centro Nacional de Investigaciones Oncológicas (CNIO)**

FECHA: febrero, 2018

DESARROLLO DE UNA BASE DE DATOS PARA LA CLASIFICACIÓN DE FÁRMACOS BASADA EN EL PROYECTO ONTOLOGÍA GÉNICA Y KEGG DRUG CON APLICACIÓN EN REPOSICIONAMIENTO DE FÁRMACOS

Autor: Raquel Benítez Dorta

Tutores: Fátima Al-Shahrour y Héctor Tejero

Ponente: Ramón Díaz Uriarte

Unidad de Bioinformática

Centro Nacional de Investigaciones Oncológicas (CNIO)

Escuela Politécnica Superior

Universidad Autónoma de Madrid

FEBRERO 2018

Índice

Resumen.....	1
Abstract.....	2
Introducción.....	3
Objetivos.....	8
Material y métodos.....	9
Clasificación.....	9
Kegg Drug: TBC.....	9
ATC.....	9
USP.....	10
GDSC.....	10
CTRP.....	10
EPC.....	10
MOA.....	11
Unificación de las clasificaciones.....	11
Análisis de enriquecimientos de conjuntos de fármacos.....	12
GSEA.....	12
GSEA prerankeado.....	13
Enriquecimiento sencillo de fármacos.....	13
Datos de sinergias.....	14
Herramienta web: DSEA.....	15
Resultados.....	16
Clasificación resultante.....	16
Análisis de enriquecimientos de conjuntos de fármacos.....	17
Enriquecimiento sencillo de fármacos.....	17
Relaciones sinérgicas entre familias de fármacos.....	18
Discusión.....	32
Conclusión.....	36
Bibliografía.....	37

Resumen

En los últimos años, el conocimiento de las peculiaridades moleculares que caracterizan el cáncer ha aumentado enormemente. En cambio, continua siendo la principal causa de muerte a nivel mundial. Reducir los costes que suponen el desarrollo de nuevas terapias es un objetivo fundamental y un reto para poder seguir avanzando. En este contexto, los métodos computacionales pueden ser de gran ayuda.

Este proyecto, se centra en desarrollar una nueva clasificación que sea útil para el objetivo de reposicionamiento computacional de fármacos, es decir, el uso de datos genómicos y algoritmos bioinformáticos para inferir nuevas indicaciones terapéuticas de fármacos ya existentes. La clasificación generada fue evaluada por análisis de enriquecimiento de conjunto de fármacos (DSEA). Primero por un método de enriquecimiento sencillo para identificar la correlación entre fármacos y familia y, posteriormente, por un método más complejo (basado en el anterior) para señalar las relaciones entre familias de fármacos.

Los resultados señalan la importancia de las características de la clasificación de base para obtener resultados útiles que puedan enfocar ensayos de reposicionamiento in vitro. Así, el reposicionamiento computacional de fármacos se perfila como una herramienta, muy rápida y económica, que podría contribuir en el desarrollo de nuevas terapias no sólo oncológicas.

Palabras clave: Reposicionamiento de fármacos, DSEA, clasificación de fármacos, enriquecimiento por familias.

Abstract

In recent years, the knowledge of the molecular peculiarities that characterize cancer has increased enormously. Despite this progress, cancer remains the leading cause of death worldwide. Reducing the costs involved in the development of new therapies is a fundamental objective and a challenge to be able to continue advancing. In this context, computational methods can be very helpful.

This project focuses on developing a new classification that is useful for the purpose of computational repositioning of drugs, that is, the use of genomic data and bioinformatic algorithms to infer new therapeutic indications of existing drugs. The generated classification was evaluated by drug set enrichment analysis (DSEA). First, by a simple enrichment method to identify correlation between drugs and family. Later, by a more complex method (based on the previous one) to indicate the relationships between drug families.

The results indicate the importance of the base classification features to obtain useful results that can focus on in vitro repositioning assays. Thus, the computational repositioning of drugs is emerging as a very fast and economical tool which could contribute to the development of new therapies not only oncological.

Key words: Drugs repositioning, DSEA, drugs classification , enrichment by families.

Introducción

El cáncer es una enfermedad compleja de etiología desconocida. Se define como un proceso de crecimiento y diseminación incontrolada de células. Designa un amplio grupo de enfermedades que pueden afectar a cualquier parte del organismo, pudiendo invadir partes adyacente al tumor o propagarse a otros órganos, proceso conocido como metástasis (OMS, 2015).

El cáncer es la principal causa de muerte a escala mundial. Según recoge el informe mundial sobre el cáncer, realizado en 2014 por la Agencia Internacional para la Investigación del Cáncer (IARC), en 2012 se registraron en todo el mundo 14 millones de nuevos casos y 8,2 millones de muertes relacionadas con la enfermedad. Además, prevé que el número de casos nuevos aumentará en un 70% en los próximos 20 años, lo cual supondría unos 22 millones. La reducción de estas cifras es un objetivo clave, por ello, en 2013 la OMS puso en marcha el Plan de Acción Global para la prevención y el Control de las Enfermedades No Transmisibles 2013-2020, cuya finalidad es reducir la mortalidad prematura causada por el cáncer, enfermedades cardiovasculares, diabetes y enfermedades respiratorias crónicas. Así, para reducir y controlar el cáncer se emplean estrategias biomédicas destinadas a la prevención de la enfermedad, la detección temprana y un tratamiento adecuado (Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, 2014; OMS, 2015).

La prevención de la enfermedad pasa por cuatro acciones fundamentales y claramente definidas: fomentar la disminución de factores de riesgo, vacunar contra el PVH y el VHB, controlar los riesgos ocupacionales y reducir la exposición a radiaciones ionizantes y no ionizantes. La detección temprana es fundamental para reducir el número de casos y consta de dos componentes: el cribado y el diagnóstico temprano. El objetivo fundamental del tratamiento es curar el cáncer o alargar los años de vida con calidad de manera considerable. Actualmente, los tratamientos más frecuentes para tratarlo son la quimioterapia, la radioterapia y la cirugía. Otros tratamientos empleados con menor frecuencia son la inmunoterapia, la terapia dirigida, la terapia hormonal, el trasplante de células madre y de médula ósea, y la medicina de precisión (OMS, 2015). Cada tipo de cáncer requiere un tratamiento específico que puede abarcar una o más modalidades ya que normalmente se usan en combinación para mejorar su eficacia.

El conocimiento y la comprensión de las peculiaridades biológicas que diferencian a las células normales de las tumorales es fundamental para poder desarrollar terapias adecuadas para el

tratamiento de la enfermedad. Así, en los últimos años, el arsenal terapéutico oncológico ha experimentado grandes avances debido al mejor conocimiento molecular y de las vías de señalización celular encargadas de procesos neoplásico. A pesar de esto, la mayoría de los tratamientos terminan en fracaso por el desarrollo de resistencias (Housman et al., 2014). Cuando un tratamiento antitumoral inicial falla, por norma general, la decisión terapéutica es tomada ignorando el mecanismo de resistencia, lo que genera tasas de respuesta pésimas. El conocimiento de dichos mecanismos específicos de resistencias adquiridas para cada tumor resulta útil a la hora generar nuevos enfoques terapéuticos que permitan contrarrestar dichas tasas ya que cada tumor adquiere un fenotipo característico propio a través de un conjunto de alteraciones genéticas y epigenéticas (Cavallo, De Giovanni, Nanni, Forni, & Lollini, 2011; Garraway & Jänne, 2012).

Por todo esto, la medicina personalizada juega un papel cada vez más importante en la prevención, el diagnóstico, el pronóstico y los tratamientos del cáncer. Esta tiene en cuenta el perfil genético del tumor y su progresión antes de asignar un tratamiento, por lo que cada paciente es tratado con el fármaco que mejor se adecua a su tipo de neoplasia (Verma, 2012).

Un factor determinante en el desarrollo de nuevas terapias o fármacos es la farmacoeconomía. El descubrimiento de nuevos fármacos y su posterior desarrollo y validación es un proceso que requiere mucho tiempo y dinero. Entre 2003 y 2011, sólo un 13% de los fármacos que iniciaron los ensayos de fase I llegaron a ser aprobados por la Administración de Fármacos y Alimentos de EEUU (FDA). Estas cifras empeoran en oncología, donde sólo el 6'7% de nuevos fármacos llegaron a ser aprobados. Además, la media desde que comienzan los ensayos clínicos hasta que el producto es comercializado es de 8,3 años (Hay, Thomas, Craighead, Economides, & Rosenthal, 2014). Se hace necesario, por tanto, la búsqueda de alternativas igualmente eficaces pero que supongan costes menores.

Una forma de disminuir la inversión podría ser el estudio del empleo de fármacos en aplicaciones distintas a las que fueron aprobados inicialmente, lo que se conoce como reposicionamiento de fármacos. Se puede dar por casualidad, en el transcurso de ensayos clínicos o de manera dirigida a partir de un estudio previo (March-Vila et al., 2017). Un ejemplo destacado es la talidomida, un fármaco indicado para las náuseas en el embarazo retirado del mercado por sus efectos teratogénicos y que actualmente se emplea para combatir numerosas enfermedades bajo estrictos controles de supervisión. Otro es el sildenafil (Viagra), durante los ensayos clínicos para indicaciones terapéuticas como la angina de pecho o la hipertensión pulmonar, se vio que era muy eficiente para tratar la disfunción eréctil. Un ejemplo más reciente es la metformina, fármaco más

prescrito para el tratamiento de la diabetes mellitus tipo 2. Se observó a partir de evidencias preclínicas que tiene un elevado potencial terapéutico como neoadyuvante en terapias contra el cáncer (Quinn, Kitagawa, Memmott, Gills, & Dennis, 2013).

La principal ventaja del reposicionamiento de fármacos radica en la reducción del tiempo y dinero invertido en los estudios de toxicidad y farmacocinética necesarios para la aprobación de la nueva indicación, puesto que existe mucha información previa (Pantziarka, Bouche, Meheus, Sukhatme, & Sukhatme, 2014). Sacar un nuevo fármaco al mercado cuesta de media 1.000 millones de dólares y se tarda de 10 a 15 años. La mayor parte de la inversión se realiza en la fase de desarrollo clínica. En estos casos, al partir de una molécula ya conocida, se puede ahorrar millones de dólares evitando la fase I del desarrollo clínico puesto que ya se tienen hechos los ensayos de seguridad del compuesto objeto de reposicionamiento. Aunque siguen siendo necesarios el resto de ensayos clínicos antes de aprobar la nueva indicación, la probabilidad de éxito es mucho mayor. Además, se puede reducir de manera muy significativa el tiempo invertido (Sertkaya, Wong, Jessup, & Beleche, 2016).

A 31 de diciembre de 2015, se encontraban en situación de comercialización 16.970 formatos de medicamentos (PAM, 2016). A esta enorme cantidad de fármacos disponibles para un posible reposicionamiento hay que sumarle otros, como aquellos estudiados no comercializados, los que se encuentran en fase de desarrollo clínica o los dados de baja. En el caso de los fármacos con indicación oncológica, en 2013, sólo había 105 aprobados por la FDA frente a los 1803 que se encontraban en fase de desarrollo (Hay et al., 2014).

Así, el empleo de métodos computacionales puede resultar de gran ayuda a la hora de identificar nuevas posibilidades de reposicionamiento de fármacos (B. Chen & Butte, 2016; Mullard, 2014; Shim & Liu, 2014). Existen numerosos recursos de este tipo que se pueden aplicar para este fin: uso de firmas transcripcionales; métodos basados en redes, estructuras biológicas o ligandos (a partir de bases de datos públicas como PubChem, ChEMBL, y DrugBank), o técnicas de “docking” molecular, entre otras.

Un ejemplo concreto de reposicionamiento farmacológico computacional es el Connectivity Map (CMap) una colección de referencia exhaustiva de perfiles de expresión génica inducidos por fármaco (Li & Lu, n.d.; Peck et al., 2006; Aravind Subramanian et al., 2017). Se han desarrollado varios enfoques para aprovechar dicha información. Uno sería el uso de la expresión génica de los fármacos en el CMap calcular la similitud entre pares de fármacos generando una red que permite explorar las oportunidades de reposicionamiento de fármacos (Li & Lu, n.d.; Peck et al., 2006).

Otro sería comparar esos mismos perfiles con los de las enfermedades para identificar correlaciones que permitan suponer que un fármaco revierte el transcriptoma asociado a la enfermedad y, por tanto, aliviarla sintomática o causalmente (Hu & Agarwal, 2009; Li & Lu, n.d.).

La dificultad para analizar ingentes cantidades de datos ya surgió con los genes implicados en determinadas patologías y la necesidad de agruparlos según criterios de orden superior (rutas metabólicas en la que actúa, funcionalidad, localización celular, etc.) y generar una terminología/nomenclatura común. De entre todas las herramientas bioinformáticas disponibles para el análisis masivo de datos, los análisis de enriquecimiento de conjuntos (Set Enrichment Analysis) pueden ayudar en el descubrimiento de patrones en los datos biomédicos. El llamado Gene Set Enrichment Analysis (GSEA) es un método que identifica aquellas clases de genes que están sobre- e infrarrepresentadas en una lista de genes dada. Usa un enfoque estadístico para identificar conjuntos de genes significativamente enriquecidos o suprimidos, mostrando expresiones diferenciales entre las condiciones que cumple cada conjunto (A. Subramanian et al., 2005).

Extrapolando este método al reposicionamiento de fármacos podríamos hablar de un análisis de enriquecimiento de fármacos (DSEA, Drug Set Enrichment Analysis) que estudiaría si un grupo de fármacos está preferencialmente localizado en uno de los extremos de una lista ordenada de fármacos.

Para este fin, puede ser viable cualquier lista de fármacos ordenada en base a un determinado criterio. Por ejemplo, medidas de sinergia/antagonismo al actuar en combinación con otro fármaco, medidas de similitud de su estructura respecto a otro fármaco o ligando, o de la similitud entre las firmas de expresión inducidas por el fármaco. El conjunto de fármacos se obtiene a partir de una clasificación que divide a los mismos en grupos atendiendo a uno o varios criterios, como por ejemplo la relación entre el fármaco y su indicación terapéutica.

Existe mucha información disponible que relaciona los fármacos con sus diversas características. Sin embargo, continua siendo un reto integrarla y estructurarla adecuadamente (Sharp, 2017). Existen numerosas clasificaciones de fármacos como por ejemplo el Sistema de Clasificación Anatómica, Terapéutica, Química (ATC) (WHO, 2012) o la clasificación basada en la diana terapéutica de KEGG drug (Target-Based Kegg Drug) (Kanehisa et al., 2014). Aún así, se pueden considerar incompletas para el objetivo de reposicionamiento computacional de fármacos por ser demasiado concretas ya que algunas sólo incluyen fármacos aprobados, otras son demasiado específicas o tienen demasiada profundidad resultando conjuntos demasiado pequeños, entre otros

motivos. Esto podría deberse al sesgo producido porque las clasificaciones existentes han sido generadas para responder a necesidades concretas en un determinado campo.

Por todo esto, el objetivo de este proyecto es desarrollar una nueva clasificación de fármacos a partir de las ya existentes que se adecue al objetivo de reposicionamiento computacional de fármacos, probar la misma con métodos de enriquecimiento de datos.

Objetivos

1. Desarrollar una nueva clasificación de fármacos, a partir de las ya existentes, basada en diferentes criterios que se adecue al objetivo de reposicionamiento de fármacos.
2. Evaluación y aplicación de técnicas de enriquecimiento en el reposicionamiento de fármacos partiendo de la clasificación anterior.

Material y métodos

Clasificación

Para generar la nueva clasificación de conjuntos de fármacos se parte de 7 clasificaciones basadas en distintos criterios y procedentes de distintas fuentes: TBC, ATC, USP, GDSC, CTRP, EPC y MOA.

Kegg Drug: TBC

Es un recurso de información farmacológica integral para medicamentos aprobados en Japón, EEUU y Europa, unificados en base a su estructura química y/o componentes químicos, y asociados con dianas terapéuticas, metabolismo enzimático y otras informaciones de redes de interacciones moleculares. KEGG (Kyoto Encyclopedia of Genes and Genomes) proporciona un sistema de clasificación en el cual los fármacos son clasificados atendiendo a sus proteínas diana en KEGG DRUG: “Target-based classification system, TBC” (L. Chen et al., 2015) (<http://www.genome.jp/kegg/drug/>). Los fármacos se clasifican en cinco niveles:

1. Nueve grupos principales.
2. Subgrupo familia proteica.
3. Subgrupo proteína.
4. Subgrupo diana proteica.
5. Sustancia química.

ATC

Sistema de clasificación anatómico, químico, terapéutico (Anatomical Therapeutic Chemical, ATC), donde los principios activos se dividen en diferentes grupos atendiendo al órgano o sistema en el que actúan y a sus propiedades terapéuticas, farmacológicas y químicas (WHO, 2012). Los fármacos se clasifican en cinco niveles:

1. Catorce grupos anatómicos principales.
2. Subgrupos farmacológicos/terapéuticos.
3. Subgrupos químicos/farmacológicos/terapéuticos.

4. Subgrupos químicos/farmacológicos/terapéuticos.

5. Sustancia química.

USP

El sistema de clasificación la Farmacopea de Estados Unidos (USP DC) es un sistema independiente actualmente desarrollado por USP Healthcare Quality & Safety Expert Committee. Utiliza evidencias farmacoterapéuticas para fármacos aprobados por la FDA creando categorías y clases (“Medicare Model Guidelines Subcommittee —For Internal Informational Use Only Guiding Principles USP Medicare Model Guidelines v7.0 Medicare Model Guidelines Subcommittee,” 2016). Tres niveles:

1. Categorías: Subgrupo terapéutico.
2. Clases: Subgrupo terapéutico/farmacológico
3. Sustancia química.

GDSC

Clasificación generada por Genomics of Drug Sensitivity of Cancer, basada en dianas terapéuticas y rutas metabólicas objetivo del fármaco (<http://www.cancerrxgene.org/downloads>). Dos niveles:

1. Grupo diana terapéutica/ruta metabólicas.
2. Sustancia química.

CTRP

Clasificación generada por Cancer Therapeutic Response Portal basada en la diana terapéutica oncológica (<https://ocg.cancer.gov/programs/ctd2/data-portal>). Dos niveles:

1. Grupos diana terapéutica oncológica.
2. Sustancia química.

EPC

“Established Pharmacologic Class”. Clasificación generada por la FDA basada en la actividad farmacológica de los distintos fármacos (Office, Communication, Outreach, & Development, n.d.). (<https://open.fda.gov/drug/label/>). Dos niveles:

1. Grupos actividad farmacológica.

2. Sustancia química.

MOA

Clasificación generada por la FDA basada en el mecanismo de acción de los distintos fármacos (Office, Communication, Outreach, & Development, n.d.) (<https://open.fda.gov/drug/label/>). Dos niveles:

1. Grupos mecanismo de acción.
2. Sustancia química.

Unificación de las clasificaciones

Para juntar la clasificaciones en primer lugar se parsearon todos los ficheros para extraer sólo la información necesaria de cada clasificación y crear un formato común que se pudiera manejar con facilidad.

De cada clasificación se extrajo la información relativa a los niveles de organización, nombre de familias y los principios activos contenidos en cada nivel. Se realizó una normalización del texto para igualar las características tipográficas entre todas las clasificaciones. A continuación, aquellas familias que no contenían principio activo/sustancia química fueron eliminadas. También se suprimieron los duplicados, es decir, aquellos principios activos que tenían igual nombre y pertenecían a las mismas familias en todos los niveles de organización. Seguidamente, se generaron los conjuntos de fármacos unificando todos aquellos principios activos que compartían las mismas familias en todos sus niveles de organización. Finalmente, esta información se reorganizó de manera que para cada conjunto de fármacos se recogió su información de la siguiente manera:

Familia: nombre de la familia/conjunto de fármacos/drug set.

Fármacos: nombre de los distintos fármacos que componen una familia.

Nivel: nivel de organización dentro de la clasificación farmacológica original.

Clasificación: clasificación de origen.

Una vez generados los conjuntos se refinaron para poder unificar las diferentes clasificaciones con mayor facilidad. Para ello:

- Conjuntos con el mismo nombre y nivel que contienen los mismos fármacos se redujeron a un único conjunto (eliminación de duplicados).
- Cuando dos o más conjuntos tenían idéntico nombre y contenían los mismos fármacos, se conservaba sólo el conjunto con mayor jerarquía (el de nivel superior).
- Conjuntos que se denominaban igual que el conjunto del que hereda y que sus fármacos estaban contenidos en el conjunto del que heredaban también son eliminados.

Se restablecieron los niveles de clasificación en base a un criterio común:

→ Nivel A: anatómico. ATC A.

→ Nivel B: terapéutico. ATC B, TBC A, USP A.

→ Nivel C: farmacológico/terapéutico. ATC C, TBC B y C, USP B, EPC, MOA.

→ Nivel D: farmacológico/terapéutico/químico. ATC D, TBC D, CTRP, GDSC.

→ Fármacos: Sustancia química. Último nivel de todas las clasificaciones.

Por último, se realizó una revisión semimanual de todos los conjuntos generados unificando aquellos que resultaban redundantes para afinar la clasificación lo máximo posible.

Todo esto, se realizó en este proyecto con ayuda de scripts en el lenguaje de programación R.

Análisis de enriquecimientos de conjuntos de fármacos

GSEA

El Análisis de Enriquecimiento de Conjuntos de Genes (GSEA, *Gene Set Enrichment Analysis*) es un método que permite realizar un análisis de enriquecimiento de conjunto de genes (GSE, *Gene Set Enrichment*) mediante una variante del test de Kolmogorov-Smirnov para determinar si un conjunto de genes se encuentra diferencialmente expresado con respecto al resto de genes en un estudio caso-control (A. Subramanian et al., 2005). Ordena los genes en función de un criterio dado y estudia si la distribución de los genes pertenecientes al conjunto de genes es significativamente distinta de la distribución de los genes no pertenecientes a dicho conjunto. Existen tres elementos clave en este método:

El cálculo de un Enrichment Score (ES) que refleja el grado en el que un conjunto S está sobrerrepresentado en los extremos de la lista completa ordenada de genes L. Será tanto mayor,

cuanto más representado esté el gen S. La magnitud del incremento depende de la correlación de dicho gen con el fenotipo.

La estimación del nivel de importancia del ES (p-valor) mediante el uso de un test de permutación basado en el fenotipo empírico que conserva la compleja estructura de correlación de los datos de expresión génica. La permutación de las etiquetas de clase preserva las correlaciones gen-gen y, por lo tanto, proporciona una evaluación de significación más razonable desde el punto de vista biológico que la que se obtendría permutando los genes.

Ajuste por Test de Hipótesis Múltiple del nivel significancia estimado evaluando toda la base de datos de conjuntos de genes.

Además, el ES es normalizado (NES) para cada conjunto de genes teniendo en cuenta, de este modo, el tamaño de los conjuntos. La tasa de descubrimiento falso (FDR) da una estimación de la probabilidad de que un conjunto con un NES dado sea un falso positivo.

GSEA prerankeado

Es el GSEA que realiza el enriquecimiento del conjunto de genes frente a una lista de ordenada de genes dada por el usuario. A diferencia del GSEA “clásico”, el p-valor se estima por permutación de genes, con el resultado de que los genes se asignan aleatoriamente a los conjuntos siempre que mantengan su tamaño. Este enfoque no es del todo exacto porque ignora las correlaciones gen-gen, sobrestimando los niveles de significación, pudiendo dar lugar a falsos positivos. No obstante, puede ser útil para la generación de hipótesis (A. Subramanian et al., 2005).

Enriquecimiento sencillo de fármacos

El análisis de enriquecimiento de fármacos (DSEA, *Drug Set Enrichment Analysis*) se lleva a cabo aplicando un algoritmo de GSEA pre-rankeado pero empleando una lista ordenada de fármacos y un conjunto de fármacos en lugar de genes y conjunto de genes (Barbie et al., 2009). Estudia si un grupo de fármacos está preferencialmente localizado en uno de los extremos de la lista ordenada de fármacos. Para realizar el análisis hace falta una lista que contenga conjuntos de fármacos y otra de fármacos ordenada siguiendo algún criterio.

El p-valor evalúa en qué medida los valores de los fármacos dados son consistentemente pequeños (regulación positiva) o grandes (regulación negativa) para cada vía. Los ES positivos corresponden

a grupos de fármacos que tienden a estar preferentemente localizados en la parte superior de la lista ordenada, los ES negativos corresponden a grupos de fármacos que están preferentes en el otro extremo.

Para llevarlo a cabo en este proyecto se usa el paquete de R fgsea (Sergushichev, 2016).

Datos de sinergias

Los fármacos en combinación se pueden administrar de manera secuencial o simultánea. En el primer caso, el segundo fármaco se puede administrar tras el fracaso del primero o dar el primero para potenciar el efecto del segundo. Así, cuando dos fármacos se usan en combinación en un tratamiento estos pueden interactuar de manera positiva (sinergia) o negativa (antagonismo). En este caso, dos fármacos se comportan como sinérgicos cuando el efecto de estos en combinación es mayor que el efecto de los dos fármacos actuando por separado. En cambio, las interacciones antagonistas tienen lugar cuando el efecto de los fármacos en combinación es menor que el esperado si los estos actuaran por separado.

Con objeto de estudiar relaciones sinérgicas entre fármacos en tratamientos secuenciales, Koplev et al, 2017 trataron 2 líneas celulares (A375, melanoma maligno y PANC 1, adenocarcinoma pancreático) con 100 fármacos distintos que constituyen el Conjunto IV de Fármacos Oncológicos Aprobados (Programa de Desarrollo Terapéutico, Instituto Nacional del Cáncer). Las células fueron expuestas a fármacos de forma individual en una curva dosis-respuesta de 8 puntos. En un experimento a parte, las células también fueron expuestas a combinaciones secuenciales de fármacos donde el primer fármaco (A) fue dado en una dosis única y el segundo fármaco (B) fue dado 24 horas después en una curva dosis-respuesta de 4 puntos. Así, se obtuvieron 10.000 posibles combinaciones de fármacos, que se contabilizaron mediante tinción y posterior cuantificación de núcleos de manera que se distinguen las medidas después de la exposición individual a cada fármaco (A0, AA, B0 y BB) y las correspondientes a la combinación secuencial en ambos sentidos (AB y BA). Se modelaron las distribuciones de probabilidad de todos los parámetros dados, tanto los experimentos individuales como los secuenciales, usando el teorema de Bayes. Finalmente, la tabla recoge las medida de sinergia en los tratamientos secuenciales de dichos fármacos en ambos sentidos (Koplev et al., 2017).

Herramienta web: DSEA

Se ha realizado una aplicación web que facilita el manejo de los scripts generados para el enriquecimiento simple de fármacos (DSEA).

Cómo entrada por parte del usuario recibe una lista ordenada de fármacos con la puntuación preferida para el análisis. Además, debe elegir qué conjuntos de fármacos desea usar para el enriquecimiento. Por defecto, emplea la clasificación general creada con todos sus niveles, pero es posible elegir también cualquiera de las clasificaciones de partida. Además, se puede seleccionar el nivel de profundidad de clasificación que desea utilizar y/o sólo aquellos conjuntos que contengan un determinado fármaco o grupo de fármacos.

Cómo salida la aplicación devuelve una lista rankeada con los conjuntos de fármacos que presentan mayor y menor potencial sinérgico y sus correspondientes gráficas.

Con este fin, se usa Shiny. Es un paquete de R que facilita la construcción aplicaciones webs interactivas a partir R (“Shiny,” n.d.).

Resultados

Clasificación resultante

La nueva clasificación de fármacos generada tiene un total de 7770 fármacos únicos, recogidos en 1430 conjuntos divididos en 4 niveles:

Nivel	Nº fármacos*	Nº familias	Tamaño medio set	Clasificación de origen
A	2496	11	251,91	ATC A
B	3779	88	66,15	ATC B, TBC A, USP A
C	7137	786	29,85	ATC C, TBC B y C, USP B, EPC, MOA
D	3147	545	7,15	ATC D, TBC D, CTRP, GDSC
TOTAL	7770	1430	25,14	

Tabla 1: Datos generales de la nueva clasificación. * El número de fármacos únicos existentes en cada categoría y el total en la clasificación.

En la siguiente gráfica se muestran las intersecciones entre los distintos niveles de la clasificación:

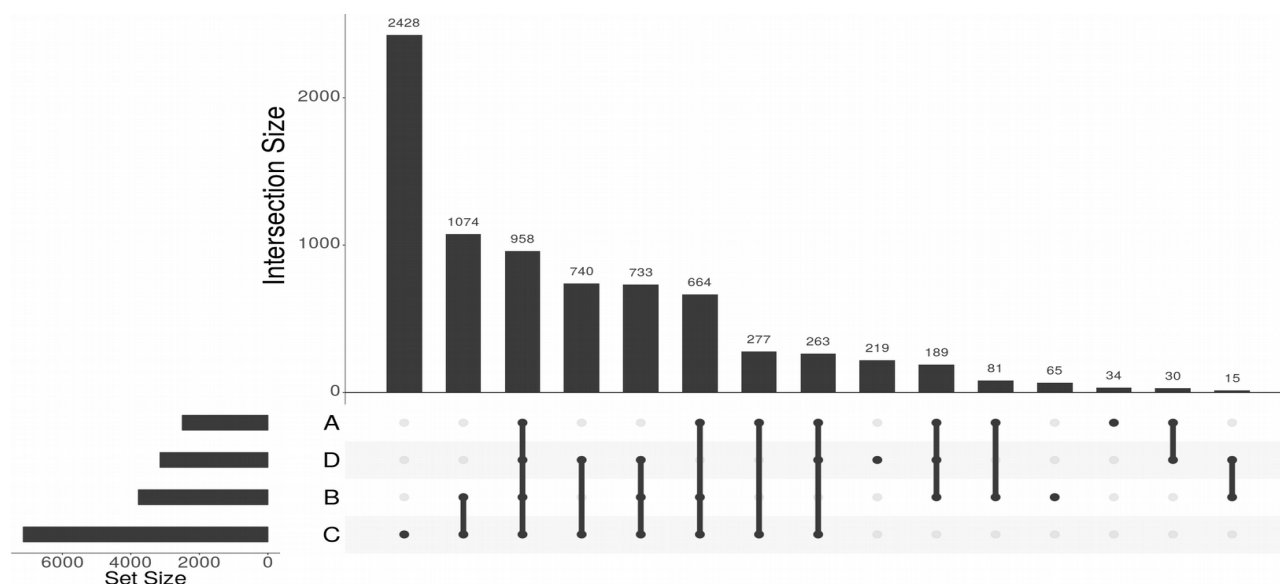


Figura 1: Diseño de una matriz para todas las intersecciones de los 4 niveles de la clasificación, ordenados por tamaño. Los puntos negros en la matriz indican los conjuntos que son parte de la intersección. “Intersection Size” es la frecuencia absoluta de intersección (el número de fármacos que sólo aparecen en la misma).

Esta clasificación generada, no es jerárquica porque no todas las clasificaciones de origen son jerárquicas. Además, las que sí lo son siguen criterios muy dispares. De esto modo, del total de fármacos diferentes recogidos en la clasificación, sólo un 13,82% se encuentra en todos los niveles y un 36,38% se encuentra en al menos 3 de los niveles.

La distribución tan desigual de fármacos se explica por el origen de las clasificaciones. La diferencia en el tamaño de los conjuntos se debe a la profundidad de cada nivel de la clasificación.

El nivel “C” se genera a partir de 6 niveles de distintas clasificaciones, por lo que presenta muchos más fármacos. En cambio, el nivel “A” tiene muchos menos fármacos porque sólo parte de 1 nivel de una clasificación. Los niveles “B” y “D” tienen como origen 3 y 4 niveles de clasificaciones distintas, respectivamente, y un número intermedio de fármacos. Así, cuanto más diversidad y cantidad de niveles de clasificación de origen se emplearon para generar un nivel, mayor cantidad de fármacos tiene dicho nivel.

Los niveles de clasificación avanzan en función de la profundidad a la hora de generar los conjuntos. “A” clasifica sus fármacos según el punto anatómico al que van dirigidos, “B” en función de su diana terapéutica, “C” atendiendo a esto y a su objetivo farmacológico, y “D” tiene en cuenta, además de lo anterior, el grupo químico al que pertenecen. Así, “A” es el nivel más general y “D” el más concreto. Esto se ve representado en el tamaño medio de los conjuntos de cada nivel, siendo más pequeños a mayor nivel de profundidad.

Análisis de enriquecimientos de conjuntos de fármacos

Enriquecimiento sencillo de fármacos

Se realizaron numerosos análisis de enriquecimiento de fármacos (DSEA) con diferentes valores de entrada. Para cada uno de ellos, se ha utilizado como entrada un conjunto de fármacos procedente de la clasificación generada en este proyecto y una lista ordena de fármacos en función de sus medidas de sinergia al ser aplicados en tratamientos secuenciales.

Se usaron los datos de Koplev et al, que recogen las medidas de sinergia resultantes de combinar 100 fármacos entre sí en dos líneas celulares. Para cada fármaco único de los 100 objeto de análisis, se realizó dos enriquecimientos por cada nivel de la clasificación, uno para el fármaco dado en primer lugar en el tratamiento secuencial y otro para cuando el fármaco es dado en segundo lugar. Fueron hechos un total de 6400 análisis de enriquecimiento sencillo de fármacos.

Nivel	A	B	C	D	ALL
Número medio familias enriquecidas (F_E)	0,81	4,96	19,4	5,16	30,20
(F_E / número familias nivel) %	7,36	5'63	2,47	0,95	2,11

Tabla 2: Media de familias enriquecidas según el nivel de la clasificación empleado en el análisis de enriquecimiento.

Así, como norma general, los resultados de los enriquecimientos que fueron mejores y más números coinciden con los realizados con el nivel C de la clasificación cuyo tamaño medio de conjunto es 29,85 y el número de fármacos únicos es 7137. Los peores resultados obtenidos fueron los correspondientes a enriquecimientos realizados con el nivel A, que sólo cuenta con 11 familias y 2496 fármacos únicos. Por ejemplo, para Nilotinib y la lista ordenada de fármacos que recoge las medidas de sinergia cuando este es dado en primer lugar en un tratamiento secuencial de la línea celular A375.

Nivel	A	B	C	D	ALL
Familias enriquecidas	1	5	19	5	29
(F_E / número familias nivel)%	9,09	5,68	2,42	0,91	2,03

Tabla 3: Total de familias obtenidas en el análisis de enriquecimiento por familias del Nilotinib cuando es administrado como primer fármaco en un tratamiento secuencial en la línea celular A375.

Sin embargo, para algunos casos concretos puede no ser útil la selección del nivel C dada la distribución desigual de fármacos y el tamaño reducido de numerosos conjuntos.

Relaciones sinérgicas entre familias de fármacos

Con el objetivo de probar los niveles de la clasificación obtenida y el método de enriquecimiento de fármacos se decidió aplicarlo a los datos de sinergia entre combinaciones secuenciales de fármacos obtenidos en Koplev et al, 2017. Se trata de encontrar combinaciones sinérgicas y antagonistas a nivel de familia de fármacos a partir de los datos a nivel de fármaco individual.

Se partió de todos los niveles de la clasificación, no se seleccionó ninguno en concreto, porque las medidas sólo corresponden a 100 fármacos únicos y, como la distribución de los principios activos es muy desigual a lo largo de la clasificación, se perderían muchos conjuntos interesantes.

Las medidas se han tomado en dos líneas celulares diferentes, A375(melanoma maligno) y PANC1(adenocarcinoma pancreático). Estas medidas varían bastante entre ellas para los mismo tratamientos secuenciales, por ello, se analizaron por separado pero siguiendo la misma metodología en ambas.

Denominaremos fármacos A son aquellos que se administran en primer lugar en el tratamiento sinérgico y fármacos B a los que administran en segundo lugar.

Primero, se hizo un primer enriquecimiento para cada fármaco A (fA) relacionado con todos los fármacos B (fB) por las medidas de sinergia. Se obtuvo la relación entre las familias de fármacos A (FA) y los fármacos B (fB). A continuación, se realizó un segundo enriquecimiento para cada familias A respecto a los fármacos B(FB). Se consiguió, así, una matriz que relaciona las familias A y las familias B (FA-FB matrix) en función de sus medidas de sinergia. Por otro lado, se repitió el proceso en sentido contrario, es decir, empezando por los fármacos B en el primer enriquecimiento y siguiendo con los A en el segundo. Se obtuvo, de este modo, otra matriz que también relaciona las familias A y las familias B, aunque se generó en orden inverso (FB-FA matrix) (Figura 2).

Se realiza en estos dos sentidos para disminuir el error producido por enriquecer en primer lugar una lista ordenada y no la otra. Es decir, el hecho de que el análisis de enriquecimiento para generar las matrices se realice teniendo en cuenta el sentido es una cuestión técnica.

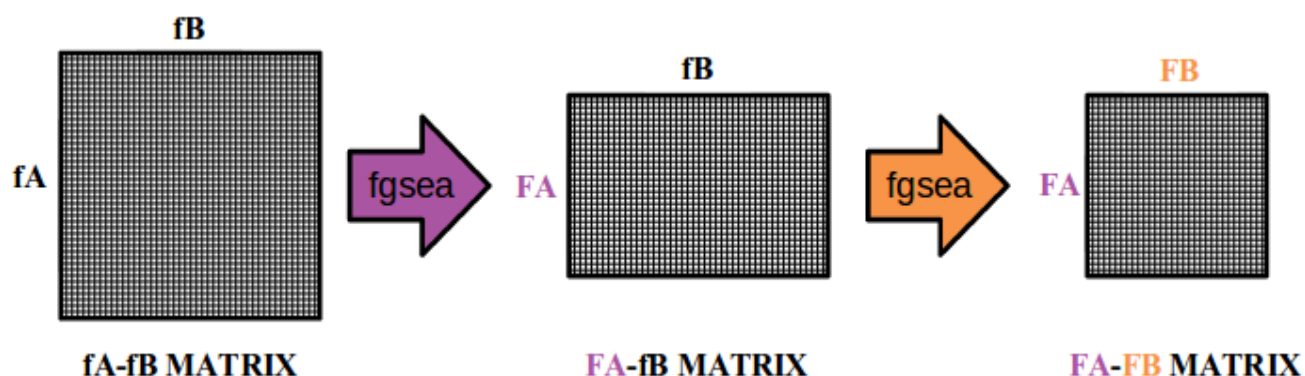


Figura 2: Esquema que describe el proceso para generar las matrices de relación sinérgica entre familias de fármacos. La matriz *fA-fB* recoge los fármacos *A* en las filas y los *B* en las columnas, cada celda corresponde a las puntuaciones de sinergia entre ellas. En la matriz *FA-fB*, las filas son familias *A*, las columnas fármacos de *B* y las celdas recogen las puntuaciones de enriquecimiento normalizadas (NES) entre las familias *A* y los fármacos *B*. La matriz *FA-FB* enfrenta las familias de fármacos *A* frente a las *B* relacionadas entre sí por su NES. Se parte de la matriz *fA-fB* y se realiza el primer enriquecimiento (*fgsea*) para obtener la matriz *FA-fB*. A continuación, se lleva a cabo el segundo enriquecimiento que genera la matriz *FA-FB*.

La siguiente tabla muestra una visión general de las 4 matrices que se generaron: dos por cada línea celular y 1 por cada posible orden de enriquecimiento posible, o primero se enriquecieron los fármacos *A* por familia y a partir de ahí los *B* o viceversa.

Número	Línea celular	Matriz	Orden sinergia	% correlación
1	A375	FA-FB	fA, fB → FA, fB → FA-FB	81,67%
2		FB-FA	fB, fA → FB, fA → FB-FA	
3	PANC1	FA-FB	fA, fB → FA, fB → FA-FB	80,60%
4		FB-FA	fB, fA → FB, fA → FB-FA	

Tabla 4: Visión general de las 4 matrices generadas.

Para comparar las matrices generadas a partir de la misma línea celular, se obtuvo el porcentaje de correlación entre dichas matrices. Este representa el grado de similitud que existe entre ambas. Se calculó contabilizando cuántas familias de las que se comportaban como sinérgicas y antagonistas

en una matriz también lo hacían en la otra. Los valores obtenidos fueron del 70,83% para las matrices 1 y 2, y del 69,42% para las matrices 3 y 4. La correlación resultó alta, pero no la esperada puesto que deberían ser prácticamente idénticas. Se sospechó, que el ruido que generaban las combinaciones aditivas aumentaban el error reduciendo la correlación, por ello se normalizaron los datos para obtener una media 0 y varianza 1. A continuación, los pares familias relacionadas entre sí por medidas de sinergia entre -1 y 1 fueron eliminadas. Tras esto, se obtuvieron los porcentajes de correlación mostrados en la tabla 4.

Teniendo en cuenta estos resultados se tomó como medida de sinergia de la interacción entre familias, la media de las interacciones FA-FB y FB-FA.

Koplev et al, 2017 observaron que las interacciones sinérgicas a nivel de fármaco eran bastante diferentes entre las dos líneas celulares estudiadas. Tras realizar la comparación entre las medidas de correlación entre las matrices que siguen el mismo orden de enriquecimiento en cada línea celular, se observó lo mismo ya que se obtuvo un porcentaje de correlación medio de 42,33%.

Para cada matriz se observó que numerosas familias antagonizan consigo mismas o con otras que contienen fármacos similares. Por ejemplo, los fármacos del conjunto que daña el DNA frente aquellos alquilantes del DNA, o los antineoplásicos con citotóxicos, independientemente del orden de administración.

Los siguientes Heat Maps representan de manera gráfica cada una de las matrices que se obtuvieron. En el eje Y, se representaron las familias de fármacos que resultaron del enriquecimiento de los fármacos que se dieron en primer lugar en el tratamiento secuencial. En el eje X, las correspondientes a los fármacos que fueron dados en segundo lugar en dicho tratamiento. Los valores individuales de cada matriz, que son puntuaciones de enriquecimientos normalizadas (NES), se representaron por colores. Los más claros se corresponden con eventos de sinergia, por el contrario, las interacciones antagonistas se representan con los colores más oscuros.

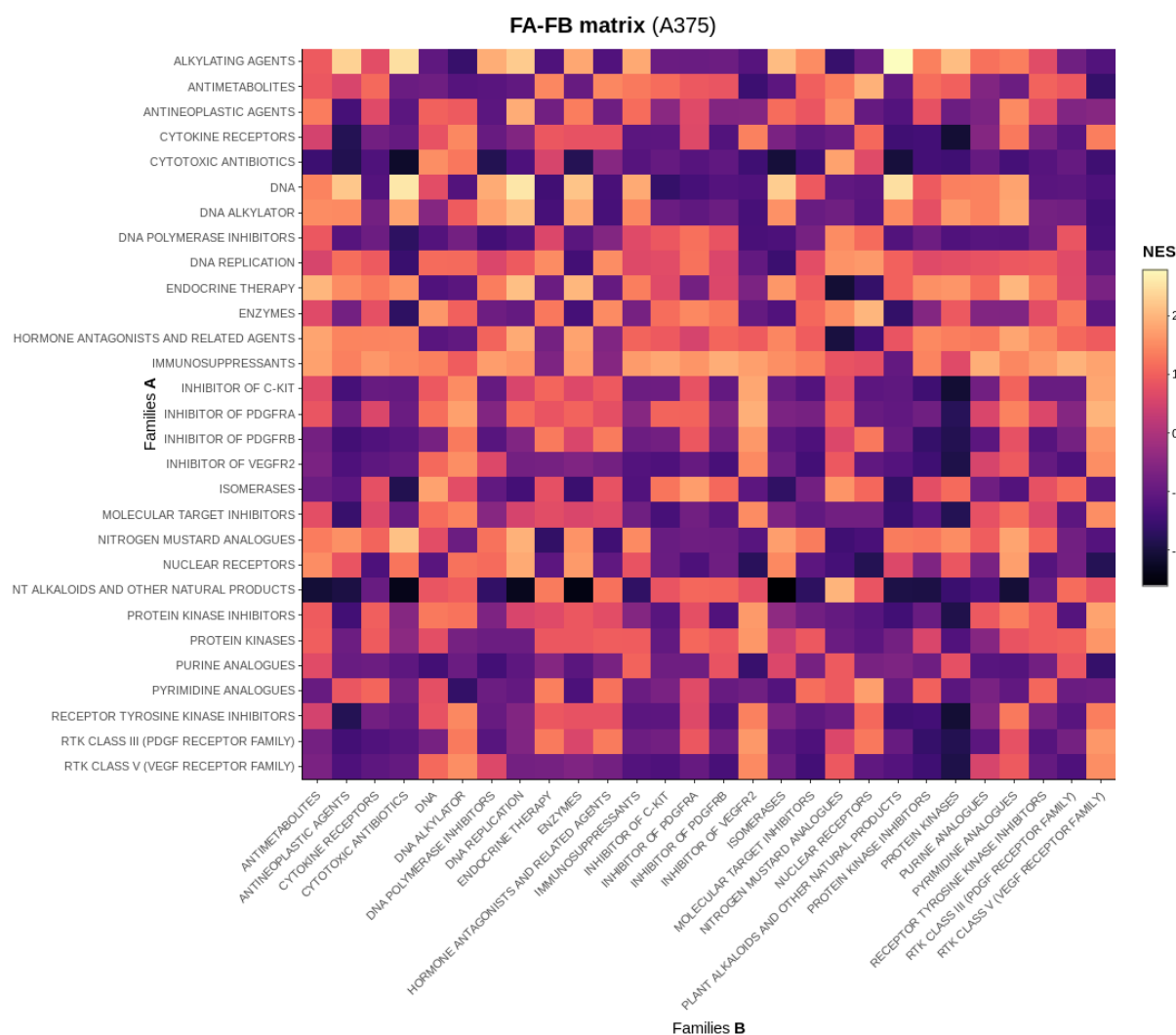


Figura 3: Heat Map correspondiente a la representación de la matriz obtenida del enriquecimiento por familias de los fármacos administrados en primer lugar (A) en el tratamiento secuencial y, a continuación, el de los fármacos dados en segundo lugar (B) a partir de los datos de la línea celular A375.

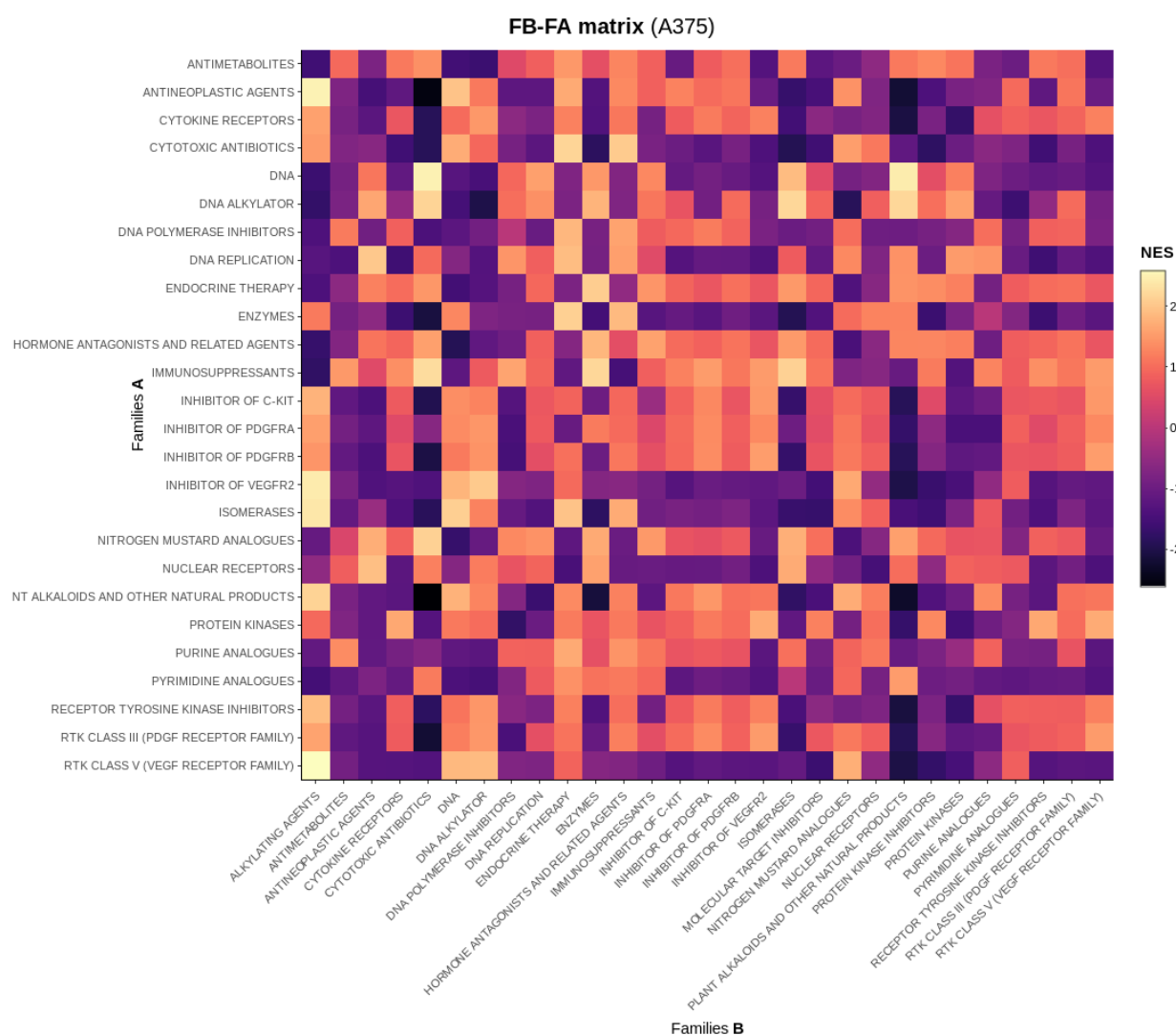


Figura 4: Heat Map correspondiente a la representación de la matriz obtenida del enriquecimiento por familias de los fármacos administrados en primer lugar (B) en el tratamiento secuencial y, a continuación, el de los fármacos dados en segundo lugar (A) a partir de los datos de la línea celular A375.

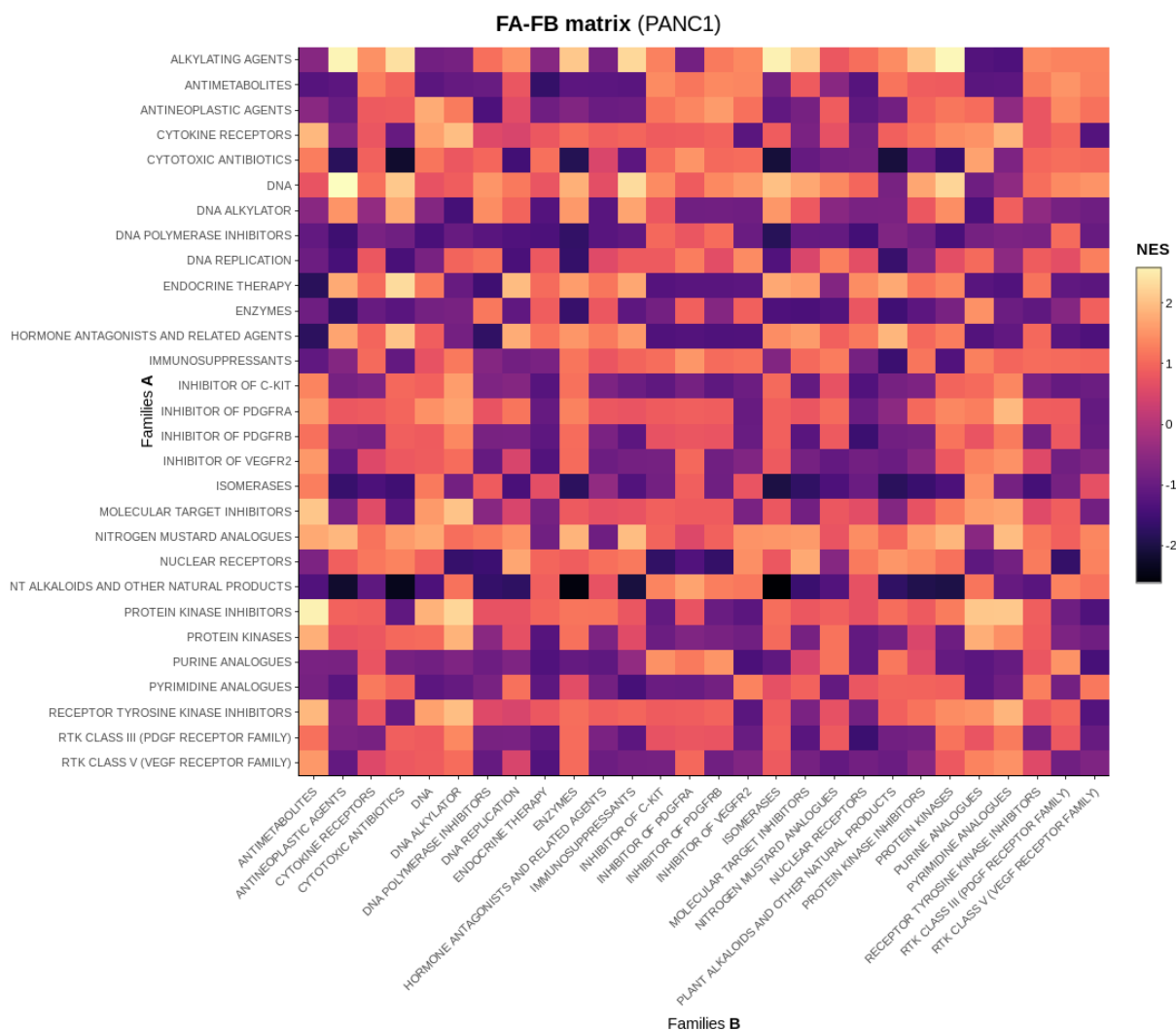


Figura 5: Heat Map correspondiente a la representación de la matriz obtenida del enriquecimiento por familias de los fármacos administrados en primer lugar (A) en el tratamiento secuencial y, a continuación, el de los fármacos dados en segundo lugar (B) a partir de los datos de la línea celular PANC1.

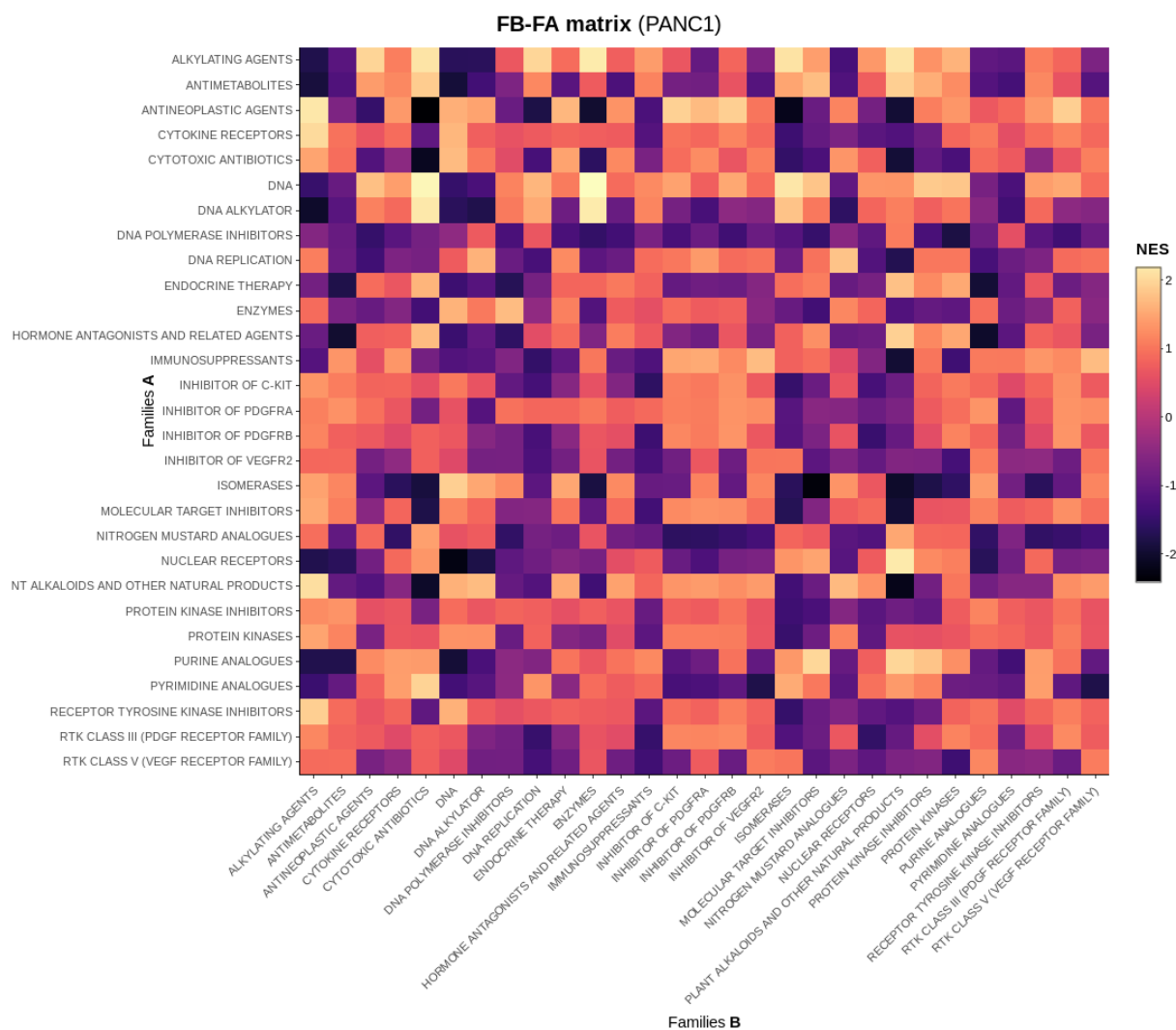


Figura 6: Heat Map correspondiente a la representación de la matriz obtenida del enriquecimiento por familias de los fármacos administrados en primer lugar (B) en el tratamiento secuencial y, a continuación, el de los fármacos dados en segundo lugar (A) a partir de los datos de la línea celular PANC1.

Podría resultar interesante el cálculo de las puntuaciones medias de sinergia (S_x) de una familia, cuando es administrada como fármaco A y cuando lo hace como fármaco B, pudiéndose dar tres escenarios:

- Un resultado próximo a 0 podría indicar que la familia no tiene efecto específico.
- En cambio, si este es mayor de 0 se podría considerar que la familia A tendería a presentar efecto sinérgico cuando actúa en esa posición del tratamiento secuencial. Por último, si el resultado fuera menor de 0, podría reflejar que la familia B tiende a un efecto antagonista cuando actúa en esa posición del tratamiento secuencial.

Con estas puntuaciones medias de sinergia, incluso, se podría analizar en qué posición del tratamiento secuencial el fármaco actúa de manera más sinérgica. Para ello, habría que comparar para una misma familia la diferencia entre sus tendencias según la posición del tratamiento secuencial en la que se dé. Esta puntuación (S_{AB}) vendría dada por la expresión:

$$S_{AB} = S_{XA} - S_{XB}$$

Dónde S_{XA} es la media de las medidas de sinergia cuando el fármaco es dado en primer lugar y S_{XB} cuando es administrado en segundo lugar. Si este resultado fuera positivo implica que la familia actúa mejor o, de manera más sinérgica, cuando se da en primer lugar. Por el contrario, si fuera negativo su acción es mejor en segundo lugar. Si esta diferencia fuera igual a 0, la posición en el tratamiento secuencial sería equivalente.

Los resultados que se han obtenido en este ejemplo se muestran en la tabla 4. Las familias de inhibidores de la ADN polimerasa, inhibidores de los RTK de clase III y los inhibidores de PDGFRB tendrían mayor efecto sinérgico empleadas en segundo lugar en el tratamiento secuencial. En cambio, las familias de agentes causantes de daños en el DNA, inmunosupresores y terapias endocrinas presentarían mayor efecto si se dan en primer lugar. Los inhibidores de tirosin quinasas no presentarían ninguna diferencia si se varía el orden de administración en el tratamiento secuencial.

Para facilitar la visualización de estas medidas, se mostraron los Heat Maps anteriores acompañados de dendogramas que ilustran las agrupaciones derivadas de la aplicación del algoritmo de enriquecimiento por familias de fármacos. Pueden ayudar a apreciar las relaciones de agrupación entre dichas familias o grupos de ellas, pero no las relaciones de similitud o cercanía dentro de estos grupos o categorías.

Families	S_{XA}	S_{XB}	S_{AB}
DNA POLYMERASE INHIBITORS	-0,4465	-0,0355	-0,411
RTK CLASS III (PDGF RECEPTOR FAMILY)	0	0,339	-0,339
INHIBITOR OF PDGFRB	0	0,3215	-0,3215
ENZYMES	-0,2145	0,0715	-0,286
ANTIMETABOLITES	-0,0355	0,1965	-0,232
INHIBITOR OF VEGFR2	0	0,2145	-0,2145
INHIBITOR OF C-KIT	0,036	0,25	-0,214
CYTOTOXIC ANTIBIOTICS	-0,0535	0,089	-0,1425
ISOMERASES	-0,161	-0,0355	-0,1255
RTK CLASS V (VEGF RECEPTOR FAMILY)	0,0175	0,125	-0,1075
PROTEIN KINASES	0,1785	0,2675	-0,089
PLANT ALKALOIDS AND OTHER NATURAL PRODUCTS	-0,1605	-0,0715	-0,089
DNA ALKYLATOR	0,107	0,143	-0,036
PURINE ANALOGUES	-0,036	-0,018	-0,018
INHIBITOR OF PDGFRA	0,1965	0,214	-0,0175
RECEPTOR TYROSINE KINASE INHIBITORS	0,25	0,25	0
CYTOKINE RECEPTORS	0,25	0,2325	0,0175
PYRIMIDINE ANALOGUES	-0,0715	-0,107	0,0355
HORMONE ANTAGONISTS AND RELATED AGENTS	0,125	0,0175	0,1075
MOLECULAR TARGET INHIBITORS	0,232	0,0715	0,1605
NUCLEAR RECEPTORS	0,0715	-0,107	0,1785
DNA REPLICATION	0,143	-0,036	0,179
ANTINEOPLASTIC AGENTS	0,1965	-0,018	0,2145
NITROGEN MUSTARD ANALOGUES	0,2675	0,036	0,2315
ENDOCRINE THERAPY	0,1785	-0,089	0,2675
IMMUNOSUPPRESSANTS	0,25	-0,0895	0,3395
DNA	0,7325	0,089	0,6435

Tabla 5: recoge las puntuaciones medias de sinergia (S_{XA} , S_{XB}) y el resultado de sus diferencias (S_{AB}).

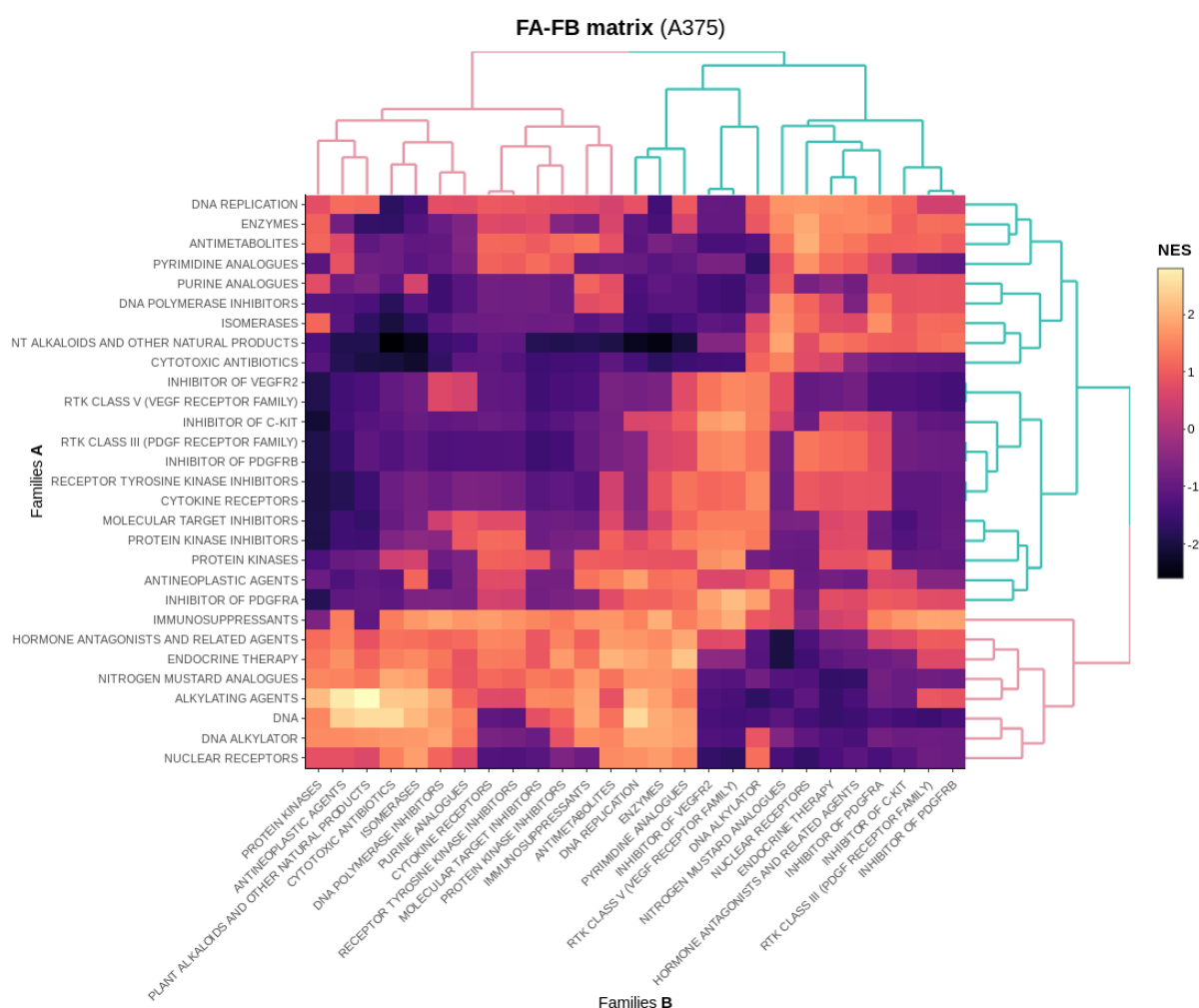


Figura 7: Heat Map y dendogramas correspondientes a la representación de la matriz obtenida del enriquecimiento por familias de los fármacos administrados en primer lugar (A) en el tratamiento secuencial y, a continuación, el de los fármacos dados en segundo lugar (B) a partir de los datos de la línea celular A375.

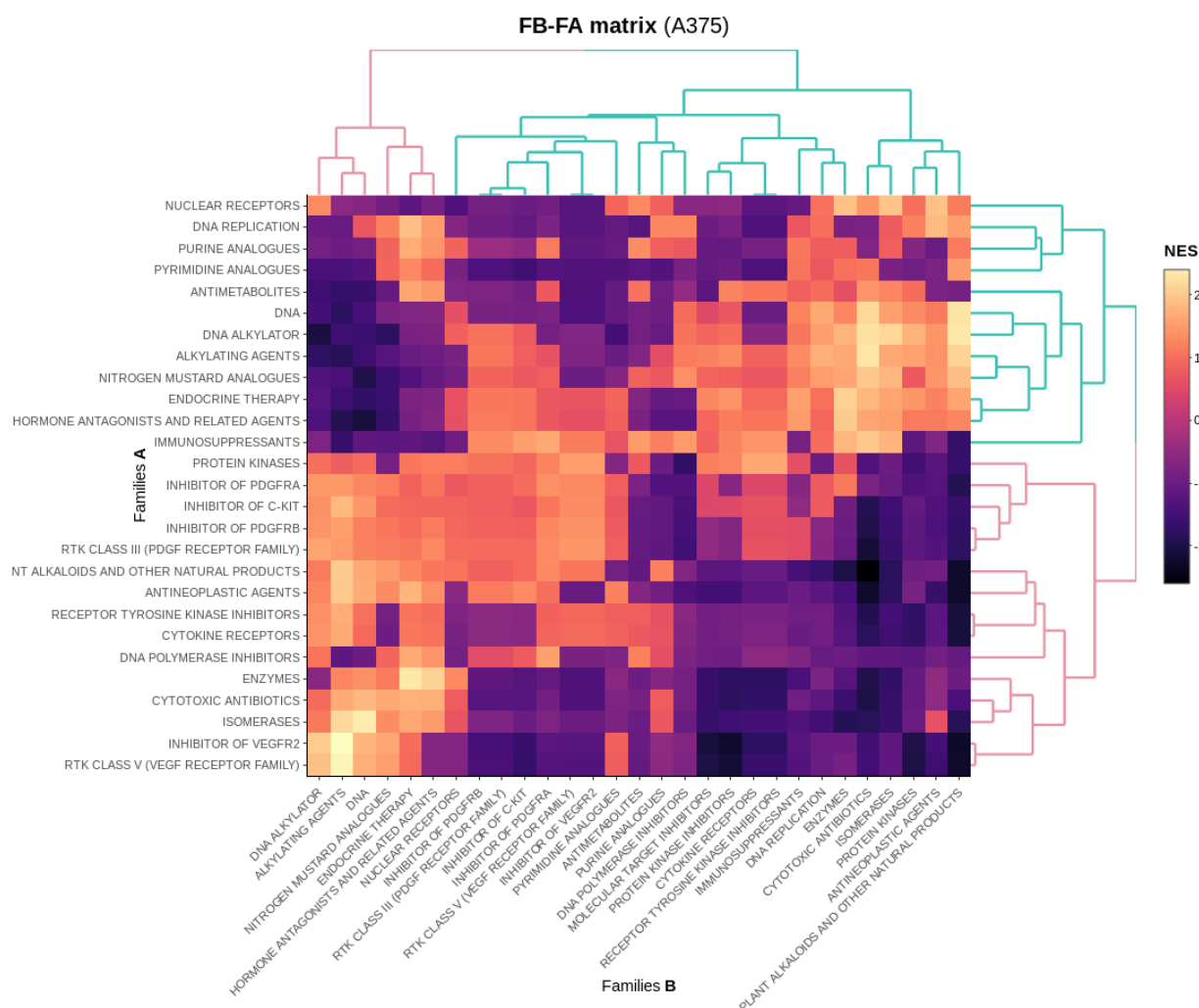


Figura 8: Heat Map y dendogramas correspondientes a la representación de la matriz obtenida del enriquecimiento por familias de los fármacos administrados en primer lugar (B) en el tratamiento secuencial y, a continuación, el de los fármacos dados en segundo lugar (A) a partir de los datos de la línea celular A375.

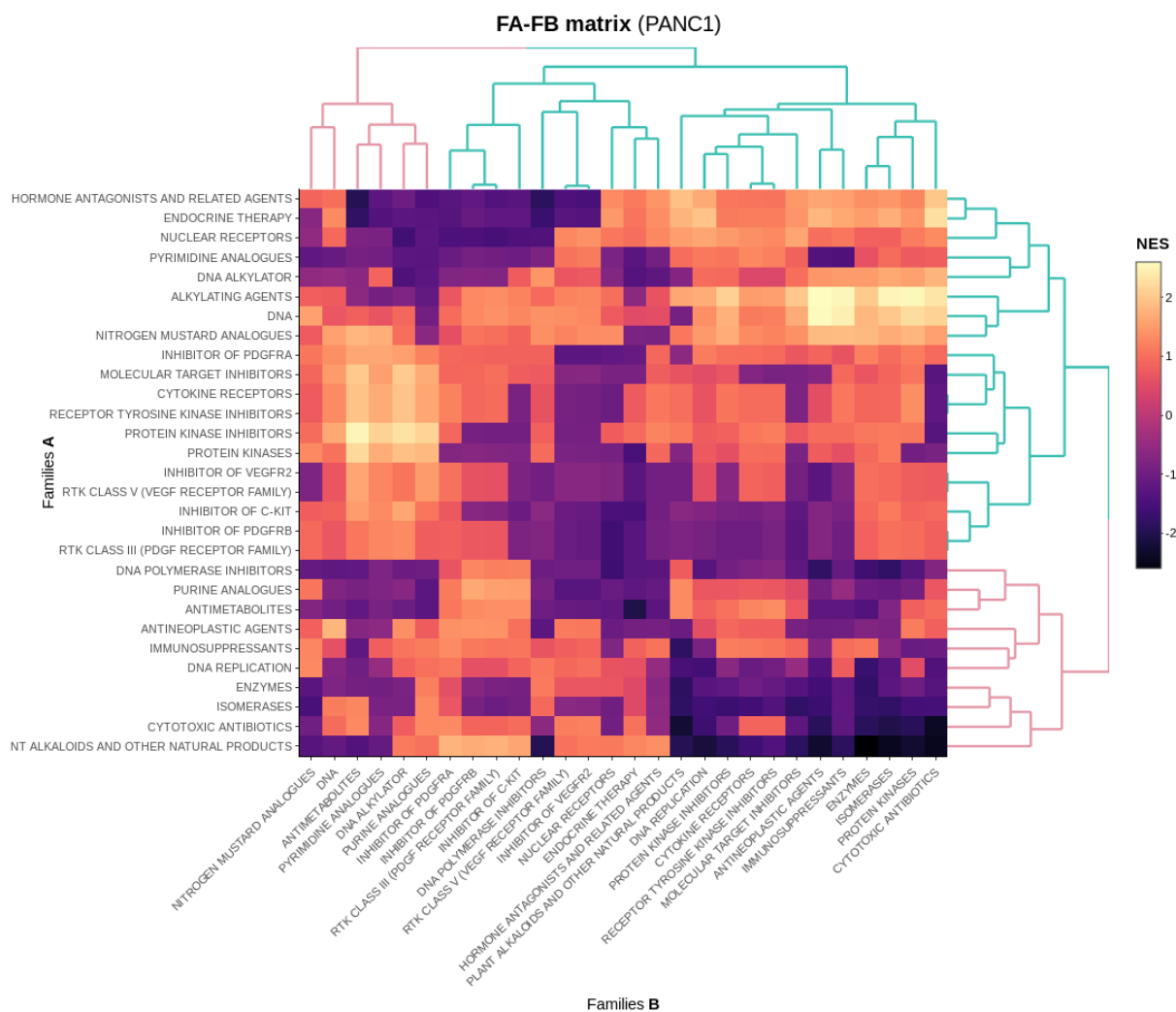


Figura 9: Heat Map y dendogramas correspondientes a la representación de la matriz obtenida del enriquecimiento por familias de los fármacos administrados en primer lugar (A) en el tratamiento secuencial y, a continuación, el de los fármacos dados en segundo lugar (B) a partir de los datos de la línea celular PANC1.

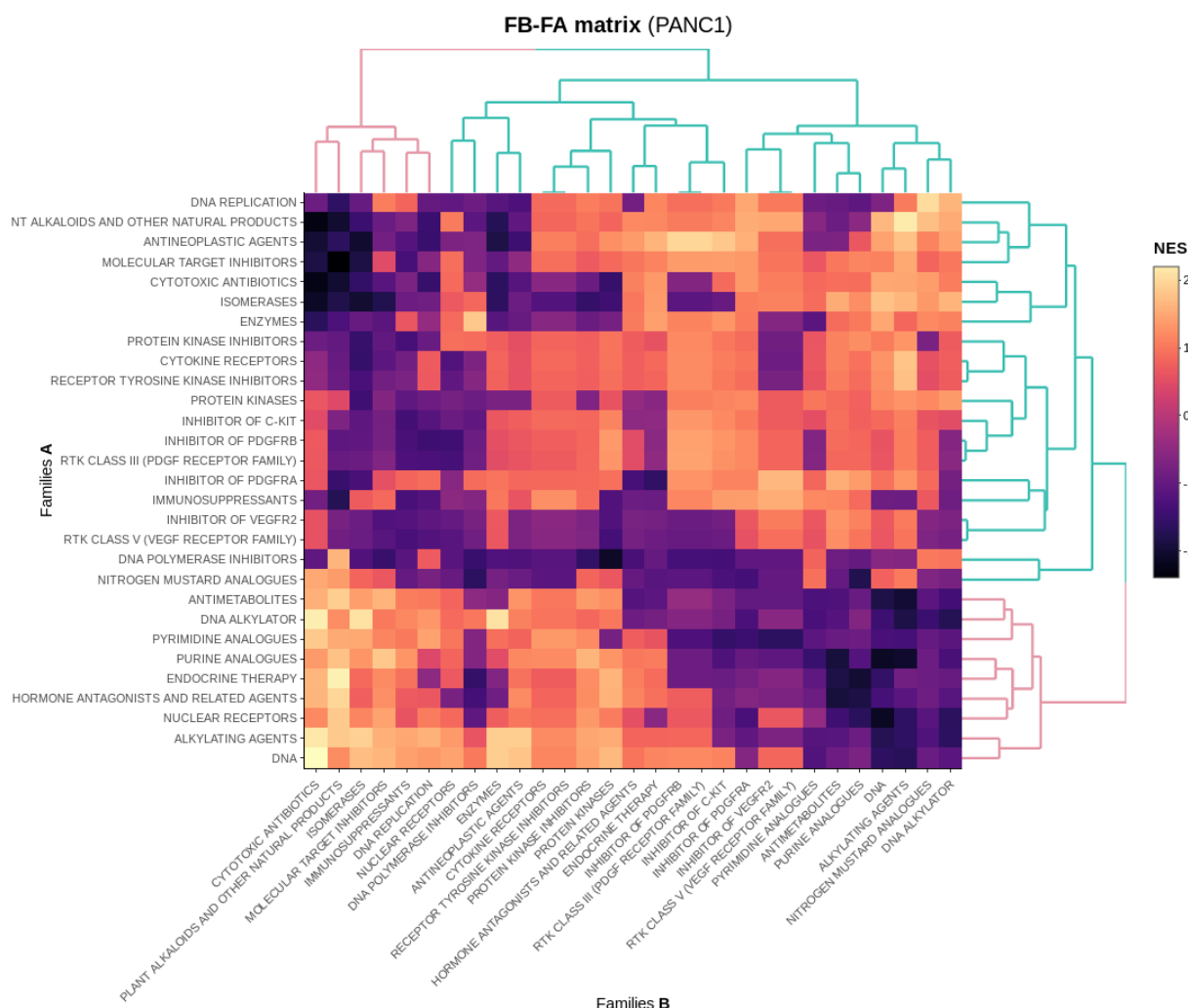


Figura 10: Heat Map y dendogramas correspondientes a la representación de la matriz obtenida del enriquecimiento por familias de los fármacos administrados en segundo lugar (B) en el tratamiento secuencial y, a continuación, el de los fármacos dados en primer lugar (A) a partir de los datos de la línea celular PANC1.

Discusión

En este proyecto, se ha desarrollado una nueva clasificación de fármacos que posteriormente se ha aplicado usando métodos de enriquecimiento con el objetivo de analizar su aplicación en el reposicionamiento computacional de fármacos.

Se trata de una clasificación de fármacos no jerárquica que integra los grupos de sustancias activas procedentes de diferentes fuentes. Con esta selección de origen, se buscó una generalización de la agrupación de fármacos. Sin embargo, en el transcurso del desarrollo de la misma surgieron numerosos problemas.

Por un lado, el nombre de un mismo fármaco no tiene porqué ser el mismo en cada clasificación ya que no existe un criterio único para denominar las sustancias activas, ni tampoco una nomenclatura común que recoja los nombres estandarizados y todos los posibles sinónimos. Esto dificulta el manejo computacional de las ingentes cantidades de información disponibles.

Por otro lado, los fármacos que incluye cada clasificación también son limitados. Ninguna de las existentes hasta el momento recoge todos los principios activos descritos, tanto aprobados como no aprobados, en fases clínicas de desarrollo o retirados del mercado. A esto hay que añadirle que cada una de estas fuentes tiene un nivel de profundidad distinto y pueden ser jerárquicas o no. De este modo, se ve limitado el proceso de integración de los datos.

Además, los fármacos se pueden agrupar de muchas maneras: por indicación, efectos secundarios, diana terapéutica, rutas metabólicas en las que intervienen, etc. Por esto, definir un criterio de generalizado y válido para un amplio campo de estudio resulta muy complejo.

En contraposición, el éxito del Proyecto de Ontología Génica (GO) como herramienta en la investigación biológica hace pensar que una adecuada ontología de fármacos podría ayudar en la mejora de las técnicas de reposicionamiento de fármacos (Hoehndorf, Dumontier, & Gkoutos, 2013). En el campo de la informática, se entiende por ontología a una definición formal de tipos, propiedades, y relaciones entre entidades que realmente o fundamentalmente existen para un dominio de discusión en particular. Podría ser una manera óptima de integrar las relaciones existentes entre los fármacos y sus muchas propiedades.

Para continuar con el desarrollado de la clasificación generada en este proyecto habría que buscar la manera de jerarquizarla, incluyendo todos los fármacos únicos en cada niveles; ampliar la colección de sustancias activas; refinar los criterios de agrupación de fármacos; conseguir que el número de sustancias activas por grupo en cada nivel sea más uniforme; incluir más niveles que incluyan otros criterios de agrupación; etc. Lo ideal sería conseguir, finalmente, una ontología a partir de toda esta información.

Dentro de todas las herramientas computacionales que podrían ser empleadas para el reposicionamiento de fármacos este proyecto se centró en el análisis de enriquecimiento de fármacos por familias. Para esto, se empleó un análisis computacional de enriquecimiento de fármacos (DSEA). Como base para este método se usó la nueva clasificación. Como datos, se tomaron los procedentes del estudio de Koplev et al, 2017 que recoge las medidas de relaciones sinérgicas entre pares de fármacos en tratamientos secuenciales.

Los resultados que se obtuvieron en el enriquecimiento sencillo de fármacos resaltaron al nivel C de la clasificación como el más apropiado para este tipo de análisis (Tabla 3). Aunque fueron mejores cuando se emplearon todos los niveles de la clasificación, aparecieron numerosas familias redundantes enriquecidas. En cambio, los resultados fueron muy reducidos para el resto de los niveles de la clasificación. Tanto el nivel B de la clasificación como el nivel D generaron una media de 5 familias enriquecidas en cada análisis. Esto se explica porque, aunque D contiene 545 familias y B sólo 88, el tamaño medio de conjunto de D (7,15) es mucho más pequeño que el de B (66,15). Todas estas diferencias, sugieren que una clasificación idónea surgiría de encontrar punto de equilibrio entre el tamaño y número de conjuntos y el número de fármacos que recoge. Así, la clasificación de base, a la hora de realizar este tipo de métodos, sería un punto crítico para la obtención de unos resultados útiles.

Con perspectiva de futuro, este método podría implementarse si, además de poder seleccionar el nivel de la clasificación que se empleará en el análisis, se pudiera escoger sólo aquellos conjuntos que contengan un determinado fármaco, o aquellos que intervengan en una determinada ruta metabólica, pudiéndolo dirigir así a campos muy diversos y concretos de investigación.

Con el objetivo de probar posibles aplicaciones para la clasificación obtenida y el análisis de enriquecimiento de fármacos se decidió aplicar los mismos a los datos de sinergia entre combinaciones secuenciales de fármacos obtenidos en Koplev et al, 2017 para encontrar combinaciones sinérgicas y antagonistas a nivel de familia de fármacos a partir de los datos a nivel

de fármaco individual. Para ello, se realizó un primer enriquecimiento sobre uno de los conjuntos de fármacos en relación al otro, obteniendo la correlación entre las familias del conjunto enriquecido y los fármacos del otro conjunto. A continuación, se realiza un segundo enriquecimiento sobre las familias ya enriquecidas en relación al otro conjunto de fármacos. Como resultado se obtuvieron dos posibles matrices de interacciones a nivel de familia dependiendo del orden en que se llevó a cabo el enriquecimiento ($A \rightarrow B$ ó $B \rightarrow A$)

La comparación entre las dos matrices de una misma línea celular genera datos de correlación en torno al 80%. Se calculó contabilizando cuántas familias de las que se comportaban como sinérgicas y antagonistas en una matriz también lo hacían en la otra. El sentido en que se generan las matrices no debería influir, puesto que se trata de una cuestión técnica. Aunque la correlación entre ellas es alta, estas tendrían que ser muy próximas al 100%. Se sigue estudiando como reducir el ruido, pero este valor estaría directamente relacionado con el número de permutaciones y el tamaño del conjunto de datos estudiados.

En cambio, al comparar las matrices generadas con este análisis (Tabla 4) para las distintas líneas celulares se observó una correlación del 42,33%. Esto hace pensar que los resultados obtenidos para cada tipo celular no son extrapolables a otros. Así, sería conveniente estudiar la selección concreta y adecuada de las familias para hacer el análisis de enriquecimiento en cada caso particular.

En todas las matrices, se encontraron ejemplos de pares de familias que recogen fármacos similares que, en cambio, mostraban comportamiento antagonista en el tratamiento secuencial independiente del orden seguido para el tratamiento secuencial. Esto podría deberse a al hecho de que poseen misma diana terapéutica o actúan sobre un mismo pathway, quedando el efecto “total” dividido entre ambas familias. Por ejemplo, en el caso de familias que dañan el ADN y los alquilantes del ADN el efecto antagonista vendría dado porque, como estos fármacos necesitan que la célula esté en fase de replicación para actuar, cuando es administrado el primer fármaco del tratamiento secuencial se detiene la misma, causando un efecto competitivo entre ambos fármacos al administrar el segundo.

El cálculo de las medias de sinergia cuando las familias son administradas en cada posición del tratamiento secuencial puede ser de gran utilidad en el estudio de los tratamientos secuenciales. Así, por ejemplo, la comparación de las mismas, además aporta información sobre la preferencia sinérgica de la familia según el orden de administración en este tipo de tratamientos.

Por ejemplo, tanto en la matriz FA-FB como en la FB-FA para ambas líneas celulares se observa la preferencia sinérgica de los inmunosupresores dados en primer lugar del tratamiento secuencial ($S_{XA} = 0,25$), frente a la preferencia antagonista si se dan en segundo lugar ($S_{XB} = -0,089$). Así, tienen mayor sinergia actuando como familia A ($S_{AB} = 0,3395$) (Tabla 5).

Las Figura 7, 8, 9 y 10 ayudan a la visualización de las preferencias sinérgicas o antagonistas de las familias de fármacos cuando son dadas como primer (A) o segundo fármaco (B) en un tratamiento secuencial. Para las familias A, cuando las filas del Heat Map son más oscuras podría indicar una preferencia antagonista, en cambio, cuando son más claras señalarían efecto preferente antagonista. Ocurre lo mismo con las columnas para las familias B.

Este método, no sólo sería válido para listas de pares de fármacos correlacionadas por sus medidas de sinergia en tratamiento secuenciales. Se podría aplicar para cualquier lista ordenada de pares de fármacos correlacionadas en base a un criterio cualquiera. Por ejemplo, sustancias activas correlacionadas entre sí según sus firmas de expresión génica, según su medidas de sus efectos secundarios, indicaciones terapéuticas, etc. Incluso, se podría utilizar con cualquier lista de pares elementos correlacionados entre sí por una puntuación. Medidas de interacción proteína- proteína para enriquecerlas por rutas metabólicas sería un ejemplo de sus posibles aplicaciones fuera del campo estrictamente farmacológico.

A pesar de esto, queda pendiente en este proyecto encontrar un manera de validar y ponderar estadísticamente la significancia de los p-valores generados con los métodos de enriquecimiento de fármacos por familia.

Conclusión

El análisis de enriquecimiento de fármacos es una herramienta con elevado potencial en la aplicación de reposicionamiento computacional del fármacos. Se trata de método que presenta un alto rendimiento ya que proporciona resultados útiles de manera rápida y económica. El avance en estos métodos permitiría ahorrar grandes cantidades de tiempo y dinero en la investigación y posterior desarrollo de nuevas terapias a partir de fármacos ya existentes.

Sin embargo, el principal factor limitante que se presenta es la necesidad del desarrollo de una forma adecuada de agrupar todos los fármacos registrados (independientemente de si están aprobados o no) según criterios generales que, posteriormente, puedan ser concretados para seleccionar las agrupaciones de interés en cada análisis individual.

Además, estas técnicas se podrían aplicar no sólo al campo farmacéutico, sino a otros muchos con interés en conocer relaciones en categorías superiores de información a partir de relaciones sencillas entre los niveles inferiores.

Por todo lo anterior, el esfuerzo y la inversión, desarrollo e innovación en técnicas computacionales aplicadas al reposicionamiento de fármacos supone una esperanza en el desarrollo de nuevas terapias.

Bibliografía

- Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., ... Hahn, W. C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*, 462(7269), 108–112. <https://doi.org/10.1038/nature08460>
- Cavallo, F., De Giovanni, C., Nanni, P., Forni, G., & Lollini, P. L. (2011). 2011: The immune hallmarks of cancer. In *Cancer Immunology, Immunotherapy* (Vol. 60, pp. 319–326). <https://doi.org/10.1007/s00262-010-0968-0>
- Chen, B., & Butte, A. J. (2016). Leveraging big data to transform target selection and drug discovery. *Clinical Pharmacology and Therapeutics*, 99(3), 285–297. <https://doi.org/10.1002/cpt.318>
- Chen, L., Chu, C., Lu, J., Kong, X., Huang, T., & Cai, Y.-D. (2015). Gene Ontology and KEGG Pathway Enrichment Analysis of a Drug Target-Based Classification System. *PloS One*, 10(5), e0126492. <https://doi.org/10.1371/journal.pone.0126492>
- Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. (2014). GLOBOCAN 2012 v1.1, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11.
- Garraway, L. A., & Jänne, P. A. (2012). Circumventing cancer drug resistance in the era of personalized medicine. *Cancer Discovery*. <https://doi.org/10.1158/2159-8290.CD-12-0012>
- Hay, M., Thomas, D. W., Craighead, J. L., Economides, C., & Rosenthal, J. (2014). Clinical development success rates for investigational drugs. *Nature Biotechnology*, 32(1), 40–51. <https://doi.org/10.1038/nbt.2786>
- Hoehndorf, R., Dumontier, M., & Gkoutos, G. V. (2013). Evaluation of research in biomedical ontologies. *Briefings in Bioinformatics*, 14(6), 696–712. <https://doi.org/10.1093/bib/bbs053>
- Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N., & Sarkar, S. (2014). Drug resistance in cancer: an overview. *Cancers*, 6(3), 1769–1792. <https://doi.org/10.3390/cancers6031769>
- Hu, G., & Agarwal, P. (2009). Human disease-drug network based on genomic expression profiles. *PLoS ONE*, 4(8). <https://doi.org/10.1371/journal.pone.0006536>
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, 42(D1). <https://doi.org/10.1093/nar/gkt1076>
- Koplev, S., Longden, J., Ferkinghoff-Borg, J., Blicher Bjerregård, M., Cox, T. R., Erler, J. T., ... Linding, R. (2017). Dynamic Rearrangement of Cell States Detected by Systematic Screening

- of Sequential Anticancer Treatments. *Cell Reports*, 20(12), 2784–2791.
<https://doi.org/10.1016/j.celrep.2017.08.095>
- Li, J., & Lu, Z. (n.d.). A New Method for Computational Drug Repositioning Using Drug Pairwise Similarity. Retrieved from
https://www.ncbi.nlm.nih.gov/CBBresearch/Lu/papers/bibm_cdr.pdf
- March-Vila, E., Pinzi, L., Sturm, N., Tinivella, A., Engkvist, O., Chen, H., & Rastelli, G. (2017). On the integration of in silico drug design methods for drug repurposing. *Frontiers in Pharmacology*, 8(MAY). <https://doi.org/10.3389/fphar.2017.00298>
- Medicare Model Guidelines Subcommittee —For Internal Informational Use Only Guiding Principles USP Medicare Model Guidelines v7.0 Medicare Model Guidelines Subcommittee. (2016). Retrieved from http://www.usp.org/sites/default/files/usp/document/our-work/healthcare-quality-safety/2016_usp_mmg_guiding_principles.pdf
- Mullard, A. (2014). 2013 FDA drug approvals. *Nature Reviews Drug Discovery*.
<https://doi.org/10.1038/nrd4239>
- Office, Communication, Outreach, & Development. (n.d.). Guidance for Industry and Review Staff Labeling for Human Prescription Drug and Biological Products — Determining Established Pharmacologic Class for Use in the Highlights of Prescribing Information Good Review Practice, 20852–21448. Retrieved from
<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>
- OMS. (2015). OMS | Cáncer. *WHO*, 297, 630. Retrieved from
<http://www.who.int/mediacentre/factsheets/fs297/es/>
- PAM. (2016). Nuevos medicamentos en 2015. *Panorama Actual Del Medicamento*, 40(390), 31–46. Retrieved from <http://pam.portalfarma.com/nuevos-medicamentos-en-2015-3/>
- Pantziarka, P., Bouche, G., Meheus, L., Sukhatme, V., & Sukhatme, V. P. (2014). The Repurposing Drugs in Oncology (ReDO) Project. *Ecancermedicalscience*.
<https://doi.org/10.3332/ecancer.2014.442>
- Peck, D., Modell, W., Blat, I. C., Crawford, E. D., Subramanian, A., Matthew, J., ... Golub, T. R. (2006). The Connectivity Map : Small Molecules , Using to Connect and Disease Signatures Genes , 313(September), 1929–1935.
- Quinn, B. J., Kitagawa, H., Memmott, R. M., Gills, J. J., & Dennis, P. A. (2013). Repositioning metformin for cancer prevention and treatment. *Trends in Endocrinology and Metabolism*.
<https://doi.org/10.1016/j.tem.2013.05.004>
- Sergushichev, A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*, 60012. <https://doi.org/10.1101/060012>

- Sertkaya, A., Wong, H. H., Jessup, A., & Beleche, T. (2016). Key cost drivers of pharmaceutical clinical trials in the United States. *Clinical Trials*, 13(2), 117–126. <https://doi.org/10.1177/1740774515625964>
- Sharp, M. E. (2017). Toward a comprehensive drug ontology: Extraction of drug-indication relations from diverse information sources. *Journal of Biomedical Semantics*, 8(1). <https://doi.org/10.1186/s13326-016-0110-0>
- Shim, J. S., & Liu, J. O. (2014). Recent advances in drug repositioning for the discovery of new anticancer drugs. *International Journal of Biological Sciences*. <https://doi.org/10.7150/ijbs.9224>
- Shiny. (n.d.). Retrieved January 24, 2018, from <https://shiny.rstudio.com/>
- Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., ... Golub, T. R. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, 171(6), 1437–1452.e17. <https://doi.org/10.1016/j.cell.2017.10.049>
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545–15550. <https://doi.org/10.1073/pnas.0506580102>
- Verma, M. (2012). Personalized medicine and cancer. *Journal of Personalized Medicine*, 2(1), 1–14. <https://doi.org/10.3390/jpm2010001>
- WHO. (2012). ATC - Structure and principles. Retrieved from http://www.whocc.no/atc/structure_and_principles/